



Europäisches
Patentamt

European
Patent Office

Office européen
des brevets

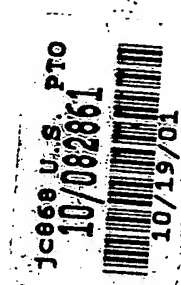
F R 000 112

4/5
5/5/2002

Bescheinigung

Certificate

Attestation



Die angehefteten Unterla-
gen stimmen mit der
ursprünglich eingereichten
Fassung der auf dem näch-
sten Blatt bezeichneten
europäischen Patentanmel-
dung überein.

The attached documents
are exact copies of the
European patent application
described on the following
page, as originally filed.

Les documents fixés à
cette attestation sont
conformes à la version
initialement déposée de
la demande de brevet
européen spécifiée à la
page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

00402938.5

**CERTIFIED COPY OF
PRIORITY DOCUMENT**

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

I.L.C. HATTEN-HECKMAN

**CERTIFIED COPY OF
PRIORITY DOCUMENT**

THIS PAGE BLANK (USPTO)



Europäisches
Patentamt

European
Patent Office

Office européen
des brevets

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.: 00402938.5
Demande n°:

Anmeldetag:
Date of filing: 24/10/00
Date de dépôt:

Anmelder:
Applicant(s):
Demandeur(s):
Koninklijke Philips Electronics N.V.
5621 BA Eindhoven
NETHERLANDS

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:
Method and device for video scene composition

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat:
State:
Pays:

Tag:
Date:
Date:

Aktenzeichen:
File no.
Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:
/

Am Anmeldetag benannte Vertragsstaaten:
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE/TR
Etats contractants désignés lors du dépôt:

Bemerkungen:
Remarks:
Remarques:

THIS PAGE BLANK (USPTO)

**CERTIFIED COPY OF
PRIORITY DOCUMENT**

FIELD OF THE INVENTION

5 The present invention relates to a method of composing a scene content from digital video data streams containing video objects, said method comprising a decoding step for generating decoded object frames from said digital video data streams, and a rendering step for composing intermediate-composed frames in a composition buffer from said decoded object frames.

10 For example, this invention may be used in the field of digital television broadcasting and implemented as an Electronic Program Guide (EPG) allowing a viewer to interact within a video rendered scene.

BACKGROUND OF THE INVENTION

15 The MPEG-4 standard, referred to as ISO/IEC 14496-2, provides functionality for multimedia data manipulation. It is dedicated to scene composition containing different natural or synthetic objects, such as two or three dimensional images, video clips, audio tracks, texts or
20 graphics. This standard allows scene content creation usable and compliant with multiple applications, allows flexibility in object combination, and offer means for user-interaction in scenes containing multiple objects. This standard may be used in a communication system comprising a server and a client terminal via a communication link. In such applications, MPEG-4 data exchanged between both sets are streamed on said communication link and
25 exploited on the client terminal to create multimedia applications.

 The international patent application WO 00/01154 describes a terminal and method of the above kind for composing and presenting MPEG-4 video programs. This terminal comprises :

- 30 - a terminal manager for managing the overall processing tasks,
 - decoders for providing decoded objects,
 - a composition engine for maintaining, updating and assembling a scene graph of the decoded objects,
 - a presentation engine for providing a scene for presentation.

SUMMARY OF THE INVENTION

35 It is an object of the invention to provide a cost-effective and optimized method of video scene composition. The invention takes the following aspects into consideration.

The composition method according to the prior art allows the composition of a video scene from a set of decoded video objects. To this end, a composition engine maintains updates and assembles a scene graph of a set of objects previously decoded by a set of decoders. In response, a presentation engine retrieves a video scene for presenting on output devices such as a video monitor. Before rendering, this method allows to individually convert decoded objects into an appropriated format. If the rendered scene format must be enlarged, a converting step must be applied to all decoded objects from which the scene is composed. This method remains then expensive since it requires high computational resources and since the complexity of threads management is increased.

To solve the limitations of the prior art method, the method of composing a scene content according to the invention is characterized in that it comprises a scaling step applied to said intermediate-composed frames for generating output frames constituting scene content.

Indeed, by performing a scaling step on intermediate-composed frames of the final scene, enlarged frames are obtained in a single processing step, which considerably reduces the computational load.

The method of scene composition according to the invention is also characterized in that said method is intended to be executed by means of a signal processor and a signal co-processor performing synchronized and parallel tasks for creating simultaneously current and future output frames from said intermediate-composed frames. Thus, the scaling step of a current intermediate-composed frame is intended to be performed by the signal co-processor while the decoding step generating decoded object frames used for the composition of the future intermediate-composed frame is intended to be performed simultaneously by the signal processor.

The use of a signal co-processor for the scaling step allows to anticipate the decoding of objects used in the composition of the future intermediate-composed frame : object frames used in the composition of the future intermediate-composed frame can ever be decoded during the composition of the current output frame. This multi-tasks method allows a high processing optimization, which leads to faster processing, what people skilled in the art will appreciate if dealing with real-time applications.

These and other aspects of the invention will be apparent from and elucidated with reference to the embodiments described hereinafter.

BRIEF DESCRIPTION OF THE DRAWINGS

The particular aspects of the invention will now be explained with reference to the embodiments described hereinafter and considered in connection with the accompanying drawings, in which identical parts or sub-steps are designated in the same manner :

5 Fig. 1 depicts a block diagram representing a terminal dedicated to a video scene composition according to the invention,

 Fig. 2 depicts processing tasks synchronization between a signal processor and a signal co-processor as used in the invention.

10

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to an improved method of composing a scene content from input video data streams encoded according to an object-oriented video standard.

15

The invention is described in the case of a video scene composed from input video streams encoded according to the MPEG-4 standard, but it would be apparent to a person skilled in the art that the scope of this invention is not limited to this specific case but can also be applied when a plurality of video streams have to be assembled, whether encoded according to the MPEG-4 standard or other object-oriented video standards.

20

Figure 1 depicts a block diagram corresponding to a video scene content composition method according to the invention. In this preferred described embodiment, the scene is composed from a background video and a foreground video, both contained in encoded video streams encoded according to the MPEG-4 standard. The method of scene composition according to the invention comprises :

25

- a decoding step 101 for decoding input MPEG-4 video streams 102 and 103, and generating decoded object frames 104 and 105, corresponding to the background and the foreground frames, respectively. There are as many decoders for generating object frames as input video streams.

30

- a rendering step 113 for composing intermediate-composed frames in a composition buffer from these previous decoded object frames. This step includes a composition sub-step of a temporary frame number i using object frame number i of the decoded background video and object frame number i of the foreground video, i varying in an increasing order between 1 and the common number of frames contained in 104 and 105. The composition order is determined

35

by the depth of each element to be rendered : the foreground video is first mapped in the composition buffer, then the foreground video is assembled in the background video in taking into consideration assembling parameters between said object frames, such as the transparency coefficient between object frames. Rendering is taking into account an user interaction 106, such as indicating the desired foreground video position compared to the

background video, said background video occupying for example the totality of the background area. Of course, other approaches can also be considered to assemble decoded object frames, as the use of the BIFS (Binary Format for Scene) containing scene graph description of object frames. The rendering step results thus in the composition of a current intermediate-composed
5 frame stored in a composition buffer from the current object frame number i referred as to 104 and the current object frame number i referred to as 105. Then, the rendering step will next compose future intermediate-composed frame number $i+1$ from the future object frame number $i+1$ of the decoded background video and the future object frame number $i+1$ of the foreground video.

10 - a scaling step 108 for enlarging the current intermediate-composed frame number i previously rendered and contained in the composition buffer, said current frame being available at the rendering output step 107. This step enlarges rendered frames 107 along horizontal and/or vertical axis so that the obtained frame 109 occupies a larger area in view of a full screen display 110. This scaling step allows to obtain a large frame format from a small frame
15 format. To this end, pixels are duplicated horizontally and vertically as many times as the scaling factor value, not only on the luminance component but also on the chrominance components. Of course, other upscaling techniques may be used such as pixels interpolation-based techniques. For example, one may consider in a preferred embodiment that intermediate-composed frames 107 are obtained from CIF (Common Intermediate Format)
20 object frames used as background, and SQCIF (Sub Quarter Common Intermediate Format) object frames used as foreground. By applying said scaling step to frames 107 with a scaling factor equal to two, the obtained frames 109 represent a QCIF overlay video format as foreground with a CCIR-601 video format as background, said CCIR-601 being required by most of displays.

25

The method according to the invention also allows to turn off the scaling step 108. Such possibility is realized with a switching step 112 avoiding any scaling operations on rendered frames 107. This switching step is commanded by an action 111 generated for
30 example by an end user who does not want to have an enlarged video format on the display 110. To this end, the user may for example interact from a mouse or a keyboard.

With the insertion of the scaling step 108 in the composition process, this invention allows to obtain a large video frame on display 110 from MPEG-4 objects of small size. As a consequence lower computational resources are required for the decoding and the rendering
35 steps, not only in terms of memory data manipulation but also in terms of CPU (Central Processing Units). This aspect of the invention avoids then processing latencies even with low processing means currently contained in consumer products since a single scaling step is performed to enlarge all object frames contained in intermediate-composed frames.

Figure 2 depicts how the composition processing steps, also called processing tasks, are synchronized when the scene composition method according to the invention is used, an horizontal time axis quantifying tasks duration. To take advantage from the complementary processing steps to be performed on MPEG-4 input video streams, the composition method is realized thanks to two types of processing done by a signal processor (SP) and a signal co-processor (SCP), said processing means being well known by people skilled in the art for performing non-extensive data manipulation tasks and extensive data manipulation tasks, respectively. The invention proposes to use these devices in such a way that composition steps of the intermediate-composed frame number $i+1$ available in 107 starts while the intermediate-composed frame number i is being composed and rendered. To this end, the whole process, managed by a tasks manager, is split in two different synchronized tasks : the decoding task and the rendering task, the decoding task being dedicated to the decoding (DEC) of input MPEG-4 object frames, and the rendering task being dedicated to the scene composition (RENDER), the scaling step (SCALE) and the presentation of the output frames to the video output (VOUT).

As an example, the intermediate-composed frame number i is composed from object frames A and B, while the intermediate-composed frame number $i+1$ is composed from object frames C and D. Explanations are given from time t_0 assuming that in such initial conditions, decoded frames A and B are available after decoding steps 201 and 202 done by the signal processor during the composition of the frame $i-1$. First, object frames A and B are rendered in a composition buffer by the rendering step 203 using signal processor resources as described previously, for generating the intermediate-composed frame number i . Then, the scaling step 204 is applied to said intermediate-composed frame number i in order to enlarge its frame format, and for generating output frame number i . Such an operation is performed by the signal co-processor so that a minimum CPU cycles is necessary compared to a same operation performed with a signal processor. Simultaneously, the beginning of the scaling operation 204 starts the decoding 205 of object frame C used in the composition of intermediate-composed frame number $i+1$. This decoding 205 is done using signal processor resources and continues until the scaling step 204 performed by the signal co-processor is finished. The scaling 204 being finished, the obtained output frame number i is presented to the video output 206 by signal processor resources to be displayed. After that the output frame number i is sent to the video out, the decoding of object frames used for the composition of intermediate-composed frame number $i+1$ is continued. Thus, the decoding step 207 is performed with signal processor resources, said step 207 corresponding to the continuation, if not finished yet during step 205, of step 205 interrupted by step 206. This step 207 follows on with decoding step 208 performed with a signal processor resources and delivering object frame D. Note that in such a solution, decoding steps are performed in a sequential order by signal processor resources.

The synchronization between decoding and rendering tasks is managed by a semaphore mechanism, said semaphore corresponding to a flag successively incremented and decremented by different processing steps. In the preferred embodiment, after each decoding loop, as it is the case after steps 201 and 202, the semaphore is set indicating to the rendering step 203 that new object frames have to be rendered. When the rendering step 203 is finished, the semaphore is reset what simultaneously launches the scaling step 204 and the decoding step 205. The scaling step is done under interruption.

To perform real-time video rendering, rendering tasks are called at a video frequency, i.e. with a period Δt equal to 1/25 second or 1/30 second according to the video standards PAL or NTSC. Using simultaneously signal processor and signal co-processor resources, the decoding of object frames C and D used for the composition of the intermediate-composed frame number $i+1$ is started during the rendering process of the output frame number i . By this way, decoded object frames are ready to be rendered when the rendering step 209 is called by the task manager. Then, the scaling step 210 is performed simultaneously with the decoding step 211, followed by the presentation step 212 to a display of the output frame number $i+1$.

A similar process starts at time $(t_0 + \Delta t)$ to render output frame number $i+1$, said frame being obtained after a scaling step applied to the intermediate-composed frame composed from object frames decoded between times t_0 and $(t_0 + \Delta t)$, i.e. during the rendering of the output frame number i .

During the scaling step of the output frame number i , a mechanism is also proposed so that the number of decoding steps is limited to a given maximum value MAX_DEC. This mechanism counts the number CUR_DEC of successive decoding performed during the scaling step generating the output frame number i , and stops the decoding when CUR_DEC reaches MAX_DEC. The decoding step goes then in an idle mode for a while, e.g. until output frame number i has been presented to a display.

Such a mechanism avoids, during the rendering of frame number i , a too high memory consumption that would cause too many successive decoding steps of object frames used in the rendering of the output frame number $i+1$.

There has been described an improved method of composing a scene content from input video data streams encoded according to the MPEG-4 video standard. This invention may also be used for scene composition from varied decoded MPEG-4 objects, such as images or binary shapes. The scaling step, dedicated for enlarging object frames, can also take different values according to the needed output frame format. The use of signal processor resources,

simultaneously performed with signal co-processor, can also be used to perform tasks different than object frames decoding, such as the analysis and the processing of user interactions.

Of course, all these aspects can take place in the present invention without deviating from the scope and the pertinence of said invention.

5

10

This invention can be implemented in several manners, such as by means of wired electronic circuits or alternatively, by means of a set of instructions stored in a computer-readable medium, said instructions replacing at least a part of said circuits and being executable under the control of a computer, a digital signal processor or a digital signal co-processor in order to carry out the same functions as fulfilled in said replaced circuits. The invention then also relates to a computer-readable medium comprising a software module that includes computer executable instructions for performing the steps, or some steps, of the above described method.

THIS PAGE BLANK (USPTO)

CLAIMS

1. Method of composing a scene content from digital video data streams containing video objects, said method comprising a decoding step for generating decoded object frames from said digital video data streams, and a rendering step for composing intermediate-composed frames in a composition buffer from said decoded object frames, **characterized in that** said method also comprises a scaling step applied to said intermediate-composed frames for generating output frames constituting scene content.
2. Method of composing a scene content as claimed in claim 1, **characterized in that** it comprises :
- a partitioning step for identifying non-extensive data manipulation steps,
 - a partitioning step for identifying extensive data manipulation steps,
- said method being intended to be executed by means of a signal processor and a signal co-processor performing synchronized and parallel processing steps for creating simultaneously current and future output frames from said intermediate-composed frames, the signal processor being dedicated to said non-extensive data manipulation steps, and the signal co-processor being dedicated to said extensive data manipulation steps.
3. Method of composing a scene content as claimed in claim 2, **characterized in that** the scaling step of a current intermediate-composed frame is intended to be performed by the signal co-processor while the decoding step generating decoded object frames used for the composition of the future intermediate-composed frame is intended to be performed simultaneously by the signal processor.
4. Method of composing a scene content as claimed in claim 3, **characterized in that** the decoding step is limited, during the scaling step, to a maximum number of decoding of object frames used for the composition of future intermediate-composed frames.
5. Device for composing a scene content from digital video data streams containing video objects, said device comprising decoding means for providing decoded object frames from said digital video data streams, and rendering means for composing intermediate-composed frames in a composition buffer from said decoded object frames, **characterized in that** said device also comprises scaling means applied to said intermediate-composed frames for generating output frames constituting scene content.
6. Device for composing a scene content as claimed in claim 5, **characterized in that** it comprises separate processing means composed by a signal processor being dedicated to non-extensive data manipulation tasks, and by a signal co-processor being dedicated to extensive data manipulation tasks, said processing means being intended to execute synchronized and

parallel calculations for creating simultaneously current and future output frames from said intermediate-composed frames.

- 5 7. Device for composing a scene content as claimed in claim 6, **characterized in that** the scaling means applied to a current intermediate-composed frame are intended to be performed by the signal co-processor while the decoding means providing decoded object frames used for the composition of the future intermediate-composed frame are intended to be performed simultaneously by the signal processor.
- 10 8. Device for composing a scene content as claimed in claim 7, **characterized in that** the decoding means are limited, during the scaling step, to a maximum number of decoding of object frames used for the composition of future intermediate-composed frames.
- 15 9. Set top box for composing a scene content from digital video data streams encoded according to the MPEG-4 standard carrying out a method as claimed in claim 1.
- 20 10. Computer program product, for a device for composing a scene content from decoded object frames, that comprises a set of instructions, which, when loaded into said device for composing, causes said device for composing to carry out the method as claimed in claim 1.

" Method and device for video scene composition "**ABSTRACT**

5 The invention relates to a cost-effective and optimized method of composing a scene content from digital video data streams containing video objects, said method comprising a decoding step for generating decoded object frames from said digital video data streams, and a rendering step for composing intermediate-composed frames in a composition buffer from said decoded object frames. The method of scene composition according to the invention comprises

10 a scaling step applied to said intermediate-composed frames for generating output frames constituting scene content. Indeed, by performing a scaling step on intermediate-composed frames of the final scene, enlarged frames are obtained in a single processing step, which considerably reduces the computational load. The use of a signal co-processor for the scaling

15 step allows to anticipate simultaneously the decoding of objects, performed by a signal processor, used in the composition of the future intermediate-composed frame.

Reference : Fig.1

Use : Video scene compositor

THIS PAGE BLANK (USP10)

1 / 2

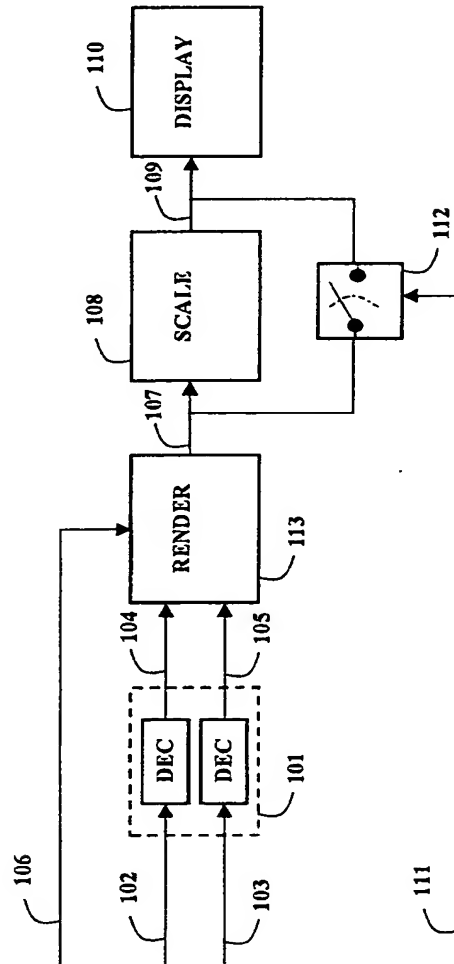


FIG.1

2 / 2

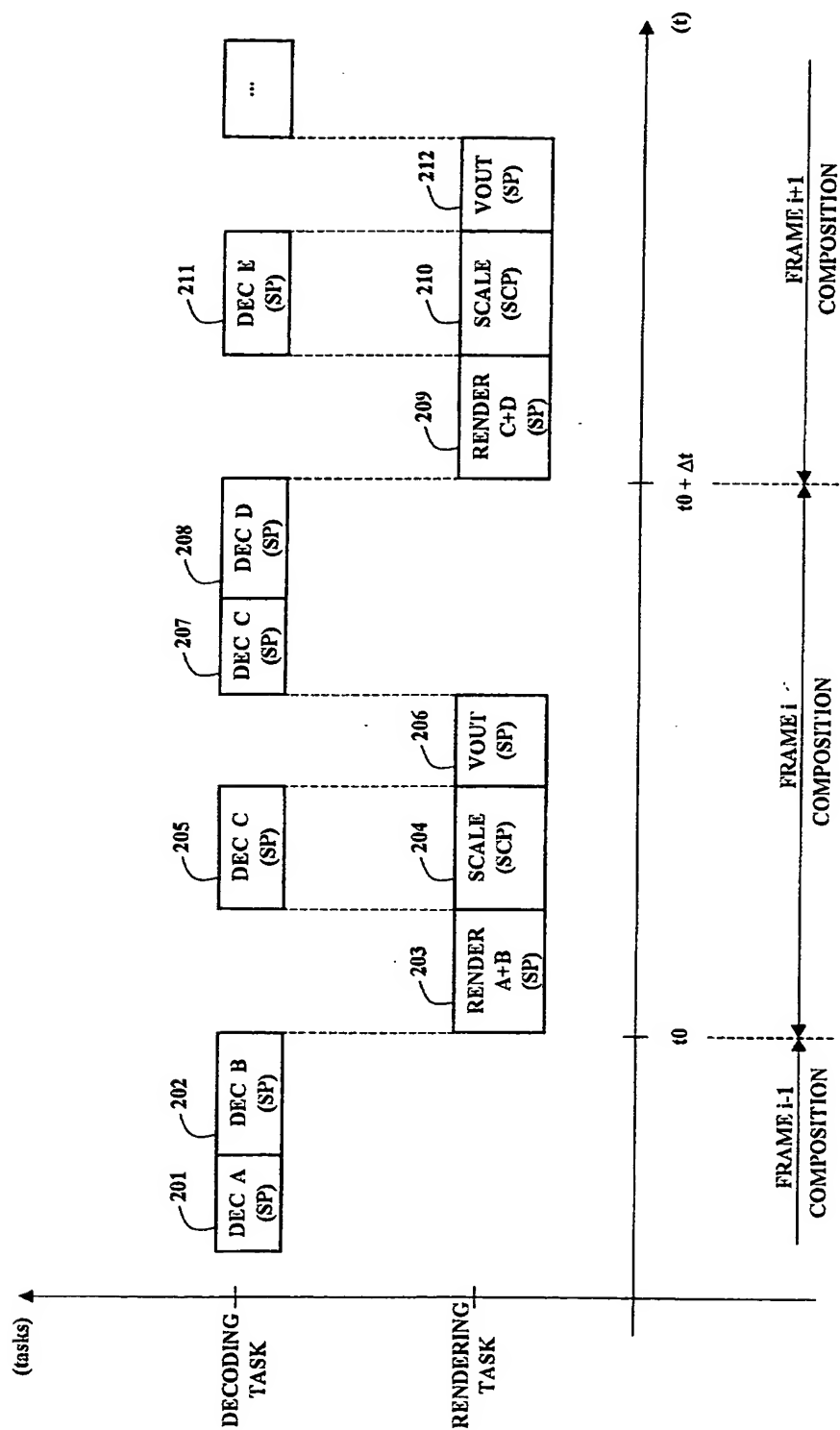


FIG.2